

Critères pour l'annotation active de microblogs

Patrick Paroubek

LIMSI - CNRS, Rue John von Neumann, Campus Universitaire d'Orsay, Bât 508, 91405 Orsay cedex
pap@limsi.fr

RÉSUMÉ

Depuis longtemps on sait optimiser l'annotation manuelle de corpus en sélectionnant les textes à annoter selon des critères rationnels plutôt qu'aléatoires, grâce aux travaux issus de l'apprentissage automatique, ceci par le biais des méthodes développées par l'apprentissage actif (« *active learning* »), un cas particulier de l'apprentissage semi-supervisé. Nous présentons ici une étude sur l'application de ces méthodes, dans le cas des contenus textuels de microblogs (Twitter) en français, annotés dans la campagne d'évaluation DEFT2015 pour des tâches d'annotation en opinions, sentiments et émotions à granularité variable. Nous comparons des critères ne supposant pas une pré-annotation, comme par exemple l'entropie de Shannon, à des critères reposant sur la quantification des divergences observées dans les pré-annotation et étudions l'impact de la quantité de données annotées déjà disponibles sur la stratégie de sélection, en prenant en compte la spécificité des contenus issus de microblogs.

ABSTRACT

Criteria for active annotation of microblogs

The methods from active learning, a particular case of semi-supervised learning, have been known for a long time as a good way to optimize the process of text selection prior to manual annotation. Rather than operate by random choice, these methods are based on rational criteria. We report here on applying such methods to French microblog text content (Twitter), annotated during the DEFT2015 evaluation campaign for several annotation tasks of opinion, sentiment or emotion mining, at different levels of semantic granularity. In this paper, we compare criteria which do not suppose the existence of a pre-annotation, like Shannon entropy, against criteria which rely on quantifying the divergences observed in pre-annotations. We study the impact that the amount of data already annotated has on the selection strategy, taking into account the specificity of microblogs content.

MOTS-CLÉS : apprentissage actif, annotation, microblogs, entropie de Shannon, évaluation, ROVER, Twitter.

KEYWORDS: Active Learning, Annotation, Microblogs, Shannon entropy, evaluation, ROVER, Twitter.

1 Introduction

Lorsque l'on contemple une tâche d'annotation de texte, quelle qu'elle soit, la question se pose de savoir si tous les contenus textuels que l'on va annoter sont indifférenciés au regard de la tâche d'annotation, ou bien si certains méritent plus d'être annotés que d'autres. La question s'est posée très tôt dans le domaine de l'**apprentissage semi-supervisé** et la réponse est venue sous la forme d'un cas particulier d'apprentissage semi-supervisé : l'**apprentissage actif** (« *active learning* »). Le

principe sur lequel repose ce dernier est l'amélioration des performances d'une tâche d'apprentissage semi-supervisé, par la présence d'un oracle parfait qui fournit au système d'apprentissage les réponses attendues exactes, ceci pour un sous-ensemble particulier des données à traiter (idéalement le plus petit possible par rapport à l'ensemble des données à traiter). Pour un état de l'art sur les travaux existants dans ce domaine, nous renvoyons le lecteur à (Settles, 2009), (Olsson, 2009) et (Maria Florina Balcan & Urner, 2015). Dans une variante de l'algorithme, l'apprentissage **proactif** (Donmez & Carbonell, 2008), l'oracle parfait est remplacé par plusieurs oracles imparfaits ayant des propriétés différentes et le problème est reformulé comme un problème de prise de décision avec incertitude, ayant pour but de trouver la paire constituée d'un oracle et d'une donnée à traiter, qui maximise l'amélioration des performances du système d'apprentissage automatique. Nous sommes alors en face d'une architecture de traitement de données comparable à celle qui est à la base de ce que l'on appelle le collaborat, ou « *crowd sourcing* » ou encore « *Human Computing* » (HC) (Law & von Ahn, 2011), sauf pour ce qui concerne la part dévolue à l'intervention humaine, part que l'on cherche à réduire dans le cas de l'apprentissage actif alors que le recours à l'humain est systématique dans le cas du HC (Bontcheva *et al.*, 2014). Par le passé (Maria Florina Balcan & Feldman, 2013) et (Berlind & Urner, 2015) ont exploré les moyens de rendre l'apprentissage actif résistant au bruit avec l'aide d'approches statistiques, tandis que (Urner *et al.*, 2013) s'est intéressé à la classification non-supervisée (« *clustering* ») et (Berlind & Urner, 2015) y ont ajouté des techniques de lissage. D'autres travaux récents se sont intéressés aux modèles à variables latentes (Acharya, 2015) pour combiner les informations provenant de différentes tâches d'annotation par le biais d'un niveau de représentation intermédiaire commun (« *abstract feature space* »). Pour ce qui nous intéresse l'apport des méthodes d'apprentissage actif concerne deux dimensions :

1. pour une donnée à traiter, choisir quel est l'oracle (ici l'annotateur) le plus approprié (Donmez & Carbonell, 2008). Au niveau d'un point de donnée unique, des approches de classification non-supervisée comme (Urner *et al.*, 2013) ou (Berlind & Urner, 2015) sont très utiles pour améliorer la performance globale de la tâche.
2. étant donné des données à traiter, comment les trier selon l'ordre inverse de leur apport potentiel d'information au modèle global de la tâche. Dans ce cadre, des mesures basées sur le taux de désaccord ou sur l'entropie informationnelle de Shannon ont été étudiées très tôt en apprentissage actif (Ayache & Quénot, 2007) (Olsson, 2009).

Parmi les travaux récents en apprentissage actif appliqués au traitement automatique des langues, nous trouvons ceux de (Sharma *et al.*, 2015) qui a adapté un échantillonnage des justifications dans une tâche de classification de texte pour résoudre les cas incertains. Quand le modèle est incertain sur l'étiquette à attribuer à un document, les mots/expressions contenus dans le document sont recherchés parmi les justifications accompagnant les documents déjà annotés. L'application de l'apprentissage actif à la compréhension de parole en environnement bruité a été abordée par (Hadian & Sameti, 2014). Quant à (Marcheggiani & Artières, 2014), ils se sont servis de quatre tâches d'étiquetage de séquence : segmentation en expressions/clauses, étiquetage morpho-syntaxique, reconnaissance d'entité nommées et reconnaissance d'entités biologiques pour comparer différentes stratégies d'apprentissage actif. La classification non-supervisée et l'apprentissage actif ont été appliqués à l'échantillonnage de mots par (Ehara *et al.*, 2014). Dans les travaux de (Dara *et al.*, 2014) ce sont des mesures d'entropie croisée qui ont été adaptées à l'apprentissage actif dans le cadre de post-edition de traductions. L'apprentissage actif a aussi été utilisé sur des corpus parallèles (anglais-chinois) avec du co-apprentissage pour l'extraction de relations sémantiques (Longhua Qian, 2014). Le déploiement conjoint d'un apprentissage actif et d'un collaborat est présenté dans (Ambati *et al.*, 2010) avec une tâche de traduction comme cas d'usage. Dans leur article, les auteurs s'intéressent aux trois contributions possibles de l'apprentissage actif pour les collaborats, à savoir, la sélection des oracles

(annotateurs), la sélection des données d'entrée et celle des données de sortie (c-a-d les informations produites par les annotateurs), ceci au niveau des phrases. (Martineau *et al.*, 2014) ont quant à eux utilisé l'apprentissage actif et le collaborat dans le cadre de fouille d'opinion sur des microblogs. Les auteurs proposent un nouvel algorithme non-linéaire de répartition de distribution (basé sur la pondération de traits Delta IDF) entre les différentes classes à apprendre afin d'identifier des traits discriminants pour les annotations incorrectes. Ces derniers travaux sont un exemple d'architecture de collaborat où les annotations erronées sont réinjectées dans le processus de traitement pour être corrigées, de manière similaire aux méthodes de « *boosting* » (Freund & Robert E. Schapire, 1999), qui peuvent être vues comme un cas particulier d'apprentissage actif appliqué aux annotations erronées.

L'apprentissage actif permet donc de disposer de stratégies améliorées de sélection/priorisation des données à annoter, les plus fréquemment utilisées reposent soit sur la prise en compte des cas les plus probables (les données pour lesquelles les prévisions sont les meilleures, c'est à dire les plus répandues ou bien les cas où l'accord entre annotations est maximal), soit au contraire, sur les cas les plus incertains (cas rares ou cas pour lesquels le taux de désaccord sur les annotations est plus élevé). Le choix entre ses deux types de stratégies dépendra à la fois de la nature de la tâche effectuée et de la quantité de données déjà annotées.

2 Cadre expérimental

Dans le cadre de la campagne d'évaluation DEFT2015 (Hamon *et al.*, 2015) qui portait sur la fouille d'opinions, de sentiments et d'émotions, un corpus de référence de messages en français issus de la plateforme Twitter¹ portant sur le changement climatique a été manuellement annoté par ELDA² avec le soutien du projet uComp³. Les annotations décrivent les opinions, sentiments ou émotions qui sont exprimées dans les messages (voir Figure 1 pour un exemple d'annotation). Les annotations sont constituées d'une part des groupes de mots qui identifient les différentes séquences de mots constituant une expression d'opinion (la source, la cible, l'expression de sentiment, les éventuels modificateurs et marqueurs de négation) et d'autre part des relations entre ces différents groupes de mots, en plus d'une annotation globale du message pour ce qui concerne la catégorie d'opinion, de sentiment ou d'émotion que le message exprime dans sa globalité, choisie parmi un ensemble de 20 catégories. Dans la campagne d'évaluation il y avait quatre tâches avec une finesse de granularité sémantique croissante :

1. déterminer la polarité des tweets (positive, négative, neutre),
2. (a) identifier la classe générale de subjectivité associée au message entier parmi : opinion, sentiment, émotion, ou information.
- (b) identifier la classe spécifique parmi 20 classes (19 classes de subjectivité et une d'information/demande)
3. identifier la source, la cible, l'expression d'opinion, de sentiment ou d'émotion, les modificateurs d'intensité et les marqueurs de négation ainsi que les relations qui les lient ensemble.

Douze équipes ont participé à la campagne d'évaluation (Hamon *et al.*, 2015) et il faut noter qu'aucune équipe n'a rendu de résultat pour la tâche 3 qui avait la granularité d'annotation la plus fine car

1. <https://twitter.com/>

2. <http://www.elda.org>

3. <http://www.ucomp.eu>

OSEE_GLOBALE valorisation

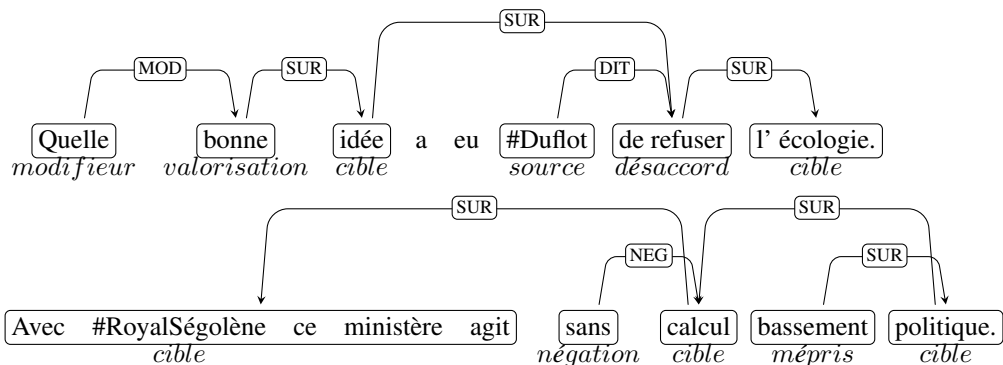


FIGURE 1 – Le sentiment global associé au Tweet (OSEE_GLOBALE) est la *valorisation*. Les relations *NEG* et *MOD* relient respectivement les marqueurs de négation (« sans ») et de modificateur d'intensité (« Quelle ») à l'expression de sentiment qu'ils modifient, ici respectivement « calcul » et « bonne ». Les liens *SUR* relie l'expression de sentiment aux objets qu'elle qualifie, notez ici que le mot « calcul » qui ne porte pas a priori de charge émotionnelle en acquiert une par le biais de la chaîne de relations *SUR* qui part du terme « bassement ». Finalement, la relation *DIT* connecte le groupe de mot qui référence le porteur d'opinion à l'expression de subjectivité.

cette dernière a été jugée trop difficile. Les annotations de références pour les différentes tâches ont été obtenues par projection à partir des annotations ayant le grain le plus fin (niveau des mots) dont un exemple a été donné dans la Figure 1. Parmi les corpus produits lors de la campagne ce sont les trois corpus obtenus en combinant les sorties des systèmes par vote majoritaire pondéré en fonction de leur performance - la mesure de précision - (Fiscus, 1997), qui vont nous intéresser. Nous allons nous poser la question de savoir s'il existe un ordre préférentiel pour annoter ces contenus de microblog que ce soit par le biais d'une annotation manuelle (par des experts ou par collaborat) ou automatique, ceci aussi bien en considérant uniquement le contenu textuel ou bien le contenu textuel enrichi d'annotations produites automatiquement. Nous regarderons en particulier l'impact de la nature des textes de microblogs sur les critères issus de l'apprentissage actif généralement employés pour effectuer cette sélection.

2.1 Choix basé sur le texte seul

Le lecteur pourra se reporter à (Neubig & Duh, 2013) et (Liao *et al.*, 2015) pour ce qui concerne les valeurs d'entropie comparatives sur des grands corpus de microblogs. Pour ce qui concerne le corpus DEFT, la Figure 2 présente les valeurs d'entropie de caractère que nous avons mesuré sur le corpus de test, avec une valeur moyenne de $E = 21.01215$ d'entropie de caractère par Tweet. Les tables 1 et 2 donnent les 10 tweets du corpus de tests de DEFT2015 ayant les valeurs d'entropie en caractères respectivement les plus faibles et les plus élevées. On voit clairement que les tweets ayant une entropie plus élevée sont les plus intéressants car ils sont plus longs et plus riches en contenu linguistique. Ils seront donc toujours à privilégier que l'on ait déjà peu ou beaucoup de données

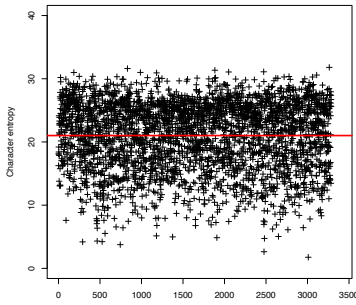


FIGURE 2 – Valeurs de l’entropie de caractère des 3285 tweets du corpus de tests DEFT2015, avec une valeur moyenne de 21.01215.

$$\begin{aligned}
 0 &< i < 3284 = N \\
 \forall m_i, 0 &< m_i < 139 \\
 t_i &= (c_0, c_1, \dots, c_{m_i}), \\
 e_i &= - \sum_{k=0}^{m_i} p(c_k) * \ln_2(p(c_k)) \quad (1) \\
 p(c_k) &= \frac{|\{c \in t_i, c=c_k\}|}{\sum_{i=0}^N m_i} \\
 E &= \frac{\sum_{i=0}^N e_i}{N}
 \end{aligned}$$

annotées.

Entropie	Tweet
1.75899	L'écosystème.
2.63123	J'AIME TROP VOIR LES ÉOLIENNES !
3.75036	LOL - - http ://t.co/GaexWGhgZc
4.18992	@afpfr super écologique.
4.22100	NETTOYAGE ÉCOLOGIQUE.wmv http ://t.co/t69dSQ44oL
4.32243	#PrénomCongolais Écosystème
4.73192	cet écosystème de pute
4.85276	Ça perturbe mon écosystème
4.98038	Ségolène Royal http ://t.co/JnRuKDGQKR
5.03911	Vélo-écosystème http ://t.co/TEs94aLRSD

TABLE 1 – Les 10 tweets du corpus DEFT 2015 ayant la plus faible entropie en caractères

2.2 Choix basé sur le texte accompagné d’annotations

Le fait de disposer d’annotation provenant de plusieurs systèmes peut aider pour la sélection des contenus à annoter/valider manuellement en considérant la divergence des annotations. La Figure 3 montre les distributions des accords d’annotation observés sur les données de tests de la campagne DEFT2015 pour les trois tâches ayant la plus grosse granularité sémantique. Sur les trois graphes de la Figure 3, nous voyons que les nombres d’accords les plus importants (impliquant un grand nombre de systèmes fournissant une réponse identique) sont ceux qui ont des nombres d’occurrences élevés (647 pour T1, 1,000 pour T2.1 et 1,366 pour T2.2). Ces valeurs sont dans la partie haute de l’intervalle de valeurs possibles pour le nombre de systèmes disant la même chose (valeur 9 sur 12 pour T1, valeur 7 sur 9 pour T2.1 et valeur 4 sur 5 pour T2.2), cf Table 3. Les valeurs d’accord correspondantes sont données dans la Table 3. Trier les textes selon l’accord maximal semble ici un bon choix lorsque

l’on dispose de peu de données annotées, car nous allons à la fois maximiser le nombre de contenus sélectionnés (partie haute de la distribution) et la confiance que nous avons dans les annotations qui

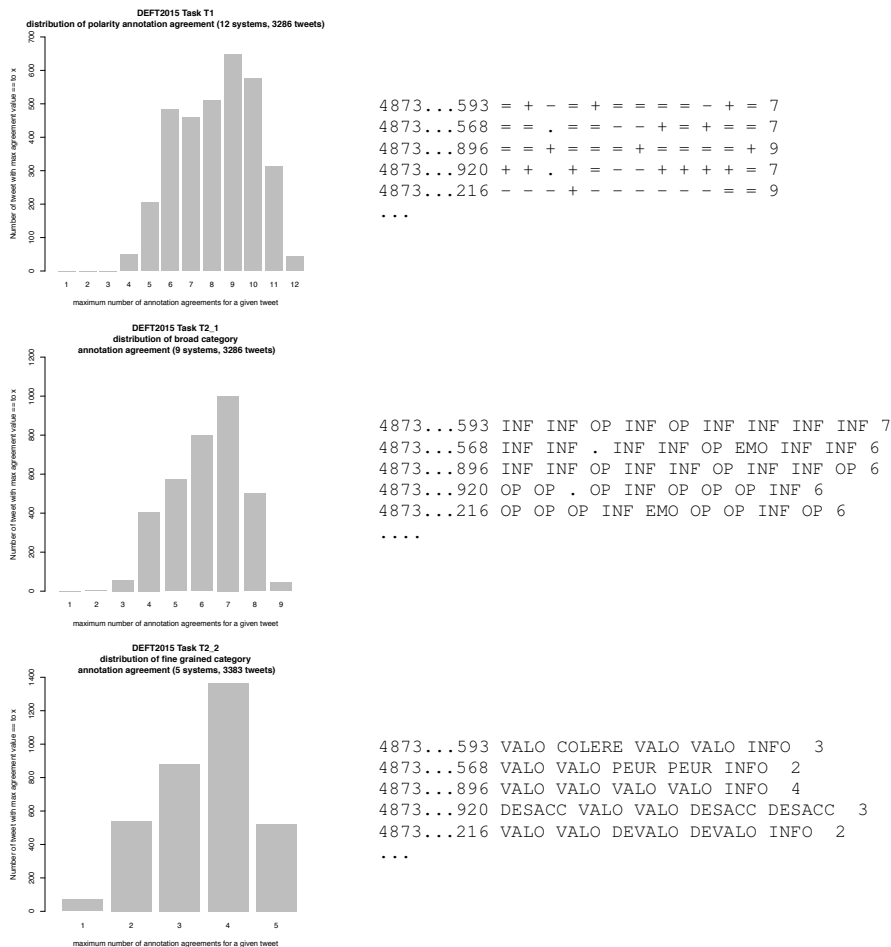


FIGURE 3 – Dans la première colonne de haut en bas, nous trouvons respectivement la distribution du nombre d’accords observés entre les systèmes par tweet pour les tâches DEFT2015 T1, T2.1 et T2.2, avec respectivement 12, 9 et 5 systèmes et 3286 tweets, 3383 tweets et 3383 tweets. La colonne de droite montre un extrait des annotation combinées pour les trois tâches.

Entropie	Tweet
30.38453	Ça me fait bien rigoler l'article qui passe sur le lait d'amande qui est de la merde écologique. On en parle de l'aberration écologique et,
30.58348	Les jeunes est un grand potentiel pour un développement durable. C'est le devoir de chacun de motiver le monde pour améliorer la jeunesse.
30.92814	Partager sa chambre avec Levantine & ; l'entendre dans son sommeil : Les éoliennes elles tournent pas parce que il fait pas assez de vent.
30.99431	Regarder une vidéo sur le réchauffement climatique. Lire les commentaires et halluciner. Certains ne croient pas que le CO2 en est une cause
31.00059	finement a quoi il sert ce parti ecologiste jamais de chez jamais je les entendu defendre ce genre d'idée ils devraient changer de nom #2
31.07053	Contribuer à la lutte contre le changement climatique, optimiser la gestion des déchets & ; offrir à ns clients des solutions éco responsables
31.09891	@Sol121Nadia @caro_blasi et electricite grace aux eoliennes off shore avec toutes les cotes et les fleuves qu on a on pourrait l etre aussi
31.34571	Je n'ai rien contre l'écologie en tant que telle. Je veux préserver la planète, comme tout le monde. Ce sont les écologistes que je déteste.
31.61390	La controverse actuelle au sein de #EELV est peut être occasion de repenser un parti de gvt avec une ligne claire : démocrate et écologiste
31.80120	Bonjour Madame le site Internet et le développement durable de faire une idée précise le monde de l'entreprise de faire une idée précise le

TABLE 2 – Les 10 tweets du corpus DEFT 2015 ayant la plus forte entropie de caractère

count task1	value task1	count task2.1	value task2.1	count task2.2	value task2.2
0	1	0	1	70	1
0	2	2	2	542	2
0	3	58	3	883	3
49	4	406	4	1,366	4
204	5	573	5	522	5
483	6	798	6	-	-
460	7	1,000	7	-	-
509	8	501	8	-	-
647	9	45	9	-	-
576	10	-	-	-	-
314	11	-	-	-	-
44	12	-	-	-	-

TABLE 3 – Distribution des accords d'annotation des participants pour les tâches T1, T2.1 et T2.2 de la campagne DEFT2015.

leurs sont associées (accord d'annotation maximal) et donc faciliter le travail des annotateurs avec une pré-annotation de bonne qualité. Sur ce point, nous pouvons améliorer la qualité des données en prenant en compte les performances des systèmes d'annotation, ce qui est le cas lorsque l'on dispose de données issues d'une campagne d'évaluation. Ici la mesure de performance qui nous intéresse est la précision car elle reflète la capacité des systèmes à fournir des réponses exactes. Ces mesures de performance serviront de poids associé au vote d'un système pour une annotation donnée (principe du ROVER (Fiscus, 1997)). On pourra se reporter à (Hamon *et al.*, 2015) pour connaître le gain obtenu avec la méthode de ROVER. La Figure 4 donne un exemple des annotation produites par combinaison pondérées pour les tâches T1, T2.2 et T2.2, avec pour chaque annotation, les différentes valeurs et la masse totale des votes pondérés qu'elle a reçue.

```

4873...896 <+, 0.835696 > <=, 5.37248 >
4873...648 <+, 0.700729 > <-, 0.699273 > <=, 4.80818 >
4873...216 <-, 4.70666 > <=, 1.50152 >
4873...593 <+, 2.00403 > <-, 0.558989 > <=, 3.64516 >
4873...568 <+, 1.29227 > <-, 1.39778 > <=, 3.51812 >

4873...648 <INFORMATION, 2.44851 > <OPINION, 0.097393 >
4873...593 <INFORMATION, 2.21321 > <OPINION, 0.332688 >
4873...920 <INFORMATION, 0.430081 > <OPINION, 2.11582 >
4873...896 <INFORMATION, 1.87631 > <OPINION, 0.66959 >
4873...216 <EMOTION, 0.332688 > <INFORMATION, 0.382598 > <OPINION, 1.83062 >

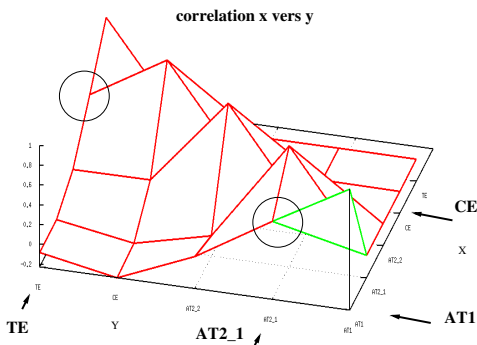
4873...896 <ENNUI, 0.0929964 > <VALORISATION, 0.45777 >
4873...568 <PEUR, 0.176961 > <VALORISATION, 0.373805 >
4873...593 <COLERE, 0.0994085 > <DERANGEMENT, 0.0929964 > <VALORISATION, 0.358362 >
4873...920 <DESACCORD, 0.358362 > <VALORISATION, 0.192405 >
4873...216 <DEVALORISATION, 0.176961 > <ENNUI, 0.0929964 > <VALORISATION, 0.280809 >

```

FIGURE 4 – Exemples de pondération obtenues pour les annotations de tweets (1 tweet par ligne) entre les différentes annotations proposées par les systèmes évalués dans les tâches T1, T2.1 et T2.2 de DEFT2015, lorsque l’on pondère les votes en fonction de la performance en précision du système ayant produit l’annotation

3 Conclusion

Pour ce qui concerne la corrélation entre différentes mesures d’accord et d’entropie nous n’avons pas observé de corrélation particulière, sauf si ce n’est un accord modéré (0.56047) d’une part pour la mesure d’accord entre systèmes pour les tâches AT1 et AT2_1 et d’autre part pour la mesure d’entropie en caractère et la mesure d’entropie en tokens définis par les espaces (0.54255). Les valeurs de corrélation sont indiquées sur le graphique de la Figure 5 et dans le tableau joint.



x/y	AT1	AT2_1	AT2_2	CE	TE
AT1	1.0	0.56047	0.09888	-0.22530	-0.07818
AT2_1	-	1.0	-0.01715	-0.20514	-0.07474
AT2_2	-	-	1.0	0.11330	0.11021
CE	-	-	-	1.0	0.54255
TE	-	-	-	-	1.0

FIGURE 5 – Les coefficients de corrélation entre les différentes mesures

Dans le cas où les contenus de microblog ne sont pas annotés, nous avons vu que le processus de sélection le plus approprié consiste à privilégier les messages ayant une entropie de caractère la plus élevée, ceci quelle que soit la quantité de données déjà annotée. Dans le cas, où l’on dispose d’annotations (qu’elles soient d’origine automatiques où non) il devient possible et intéressant de choisir les messages pour lesquels l’accord d’annotation est maximal, si de plus on dispose de mesures de performance en précision pour les systèmes d’annotation, ces mesures peuvent être utilisées pour

pondérer la combinaison des annotations, là encore les messages recevant la plus grande masse de votes seront à sélectionner en premier. Pour l’avenir, il serait intéressant d’étudier l’impact d’une segmentation en mots plutôt qu’en tokens et de considérer des mesures d’entropie alternatives.

Remerciements

Les travaux présentés ici ont été effectués dans le cadre du projet CHIST-ERA uComp ANR-12-CHRI-0003-03.

Références

- ACHARYA A. (2015). *Knowledge Transfer Using Latent Variable Models*. PhD thesis, Department of Electrical and Computer Engineering, The University of Texas at Austin.
- AMBATI V., VOGEL S. & CARBONELL J. (2010). Active learning and crowd-sourcing for machine translation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Valletta, Malta.
- AYACHE S. & QUÉNOT G. (2007). Evaluation of active learning strategies for video indexing. In *Proceedings of the 1th IEEE International Workshop on Content-Based Multimedia Indexing (CBMI’07)*, Bordeaux, France.
- BERLIND C. & URNER R. (2015). Active nearest neighbors in changing environments. *Journal of Machine Learning Research Workshop and Conference Proceedings*, **37**.
- BONTCHEVA K., ROBERTS I., DERCZYNSKI L. & ROUT D. (2014). The gate crowdsourcing plugin : Crowdsourcing annotated corpora made easy. In *Proceedings of the European Association for Computational Linguistics*, Gothenburg, Sweden : ACL.
- DARA A. A., VAN GENABITH J., LIU Q., JUDGE J. & TORAL A. (2014). Active learning for post-editing based incrementally retrained mt. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2 : Short Papers*, p. 185–189, Gothenburg, Sweden : Association for Computational Linguistics.
- DONMEZ P. & CARBONELL J. G. (2008). Proactive learning : Cost-sensitive active learning with multiple imperfect oracles. In *Proceedings of the 17th Conference on Information and Knowledge Management (CIKM’08)*, Napa Valley, California, USA : ACM.
- EHARA Y., MIYAO Y., OIWA H., SATO I. & NAKAGAWA H. (2014). Formalizing word sampling for vocabulary prediction as graph-based active learning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 1374–1384, Doha, Qatar : Association for Computational Linguistics.
- FISCUS J. G. (1997). A post-processing system to yield reduced word error rates : recognizer output voting error reduction (rover). In *In proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, p. 347–357, Santa Barbara, CA.
- FREUND Y. & ROBERT E. SCHAPIRE (1999). A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, **14**(5), 771–780.

- HADIAN H. & SAMETI H. (2014). Active learning in noisy conditions for spoken language understanding. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics : Technical Papers*, p. 1081–1090, Dublin, Ireland : Dublin City University and Association for Computational Linguistics.
- HAMON T., FRAISSE A., PAROUBEK P., ZWEIGENBAUM P. & GROUIN C. (2015). Analyse des émotions, sentiments et opinions exprimés dans les tweets : présentation et résultats de l'édition 2015 du défi fouille de texte. In *Actes de l'atelier de clôture du Xème Défi Fouille de Texte, joint à la 22ème conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, p. 1–11, Caen.
- LAW E. & VON AHN L. (2011). Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **5**(3), 1–121.
- LIAO H., FU K. & HALE S. A. (2015). How much is said in a microblog ? A multilingual inquiry based on weibo and twitter. *CoRR*, **abs/1506.00572**.
- LONGHUA QIAN, HAOTIAN HUI Y. H. G. Z. Q. Z. (2014). Bilingual active learning for relation classification via pseudo parallel corpora. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 582–592, Baltimore, Maryland : Association for Computational Linguistics.
- MARCHEGGIANI D. & ARTIÈRES T. (2014). An experimental comparison of active learning strategies for partially labeled sequences. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 898–906, Doha, Qatar : Association for Computational Linguistics.
- MARTINEAU J., CHEN L., CHENG D. & SHETH A. (2014). Active learning with efficient feature weighting methods for improving data quality and classification accuracy. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 1104–1112, Baltimore, Maryland : Association for Computational Linguistics.
- MARIA FLORINA BALCAN & FELDMAN V. (2013). Statistical active learning algorithms for noise tolerance and differential privacy. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS)*, Stateline, Nevada.
- MARIA FLORINA BALCAN & URNER R. (2015). *Encyclopedia of Algorithms*, chapter Active Learning - Modern Learning Theory. Springer.
- NEUBIG G. & DUH K. (2013). How much is said in a tweet ? a multilingual, information-theoretic perspective. In *Proceedings of the AAI Spring Symposium / Analyzing Microtext*, Palo Alto, California.
- OLSSON F. (2009). *A literature survey of active machine learning in the context of natural language processing*. Rapport interne T2009 :06, Swedish Institute of Computer Science.
- SETTLES B. (2009). *Active Learning Literature Survey*. Rapport interne 1648, University of Wisconsin–Madison.
- SHARMA M., ZHUANG D. & BILGIC M. (2015). Active learning with rationales for text classification. In *North American Chapter of the Association for Computational Linguistics – Human Language Technologies*.
- URNER R., WULFF S. & BEN-DAVID S. (2013). Plal : Cluster-based active learning. *Journal of Machine Learning Research Workshop and Conference Proceedings*, **30**.